

EFFICIENT AND ACCURATE ALGORITHMS FOR THE COMPUTATION AND INVERSION OF THE INCOMPLETE GAMMA FUNCTION RATIOS *

AMPARO GIL[†], JAVIER SEGURA[‡], AND NICO M. TEMME[§]

Abstract. Algorithms for the numerical evaluation of the incomplete gamma function ratios $P(a, x) = \gamma(a, x)/\Gamma(a)$ and $Q(a, x) = \Gamma(a, x)/\Gamma(a)$ are described for positive values of a and x . Also, inversion methods are given for solving the equations $P(a, x) = p$, $Q(a, x) = q$, with $0 < p, q < 1$. Both the direct computation and the inversion of the incomplete gamma function ratios are used in many problems in statistics and applied probability. The analytical approach from earlier literature is summarized and new initial estimates are derived for starting the inversion algorithms. The performance of the associated software to our algorithms (the Fortran 90 module **IncgamFI**) is analyzed and compared with earlier published algorithms.

Key words. Incomplete gamma function ratios, chi-squared distribution function, inversion of incomplete gamma functions, numerical evaluation of special functions, asymptotic analysis.

AMS subject classifications. 33B20, 41A60, 65D20

1. Introduction. As it is well known, the chi-squared distribution (or its equivalent, the incomplete gamma integral) plays a key role in many applied probability problems. The incomplete gamma functions are defined by

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt, \quad \Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt, \quad (1.1)$$

with ratios

$$P(a, x) = \frac{1}{\Gamma(a)} \gamma(a, x), \quad Q(a, x) = \frac{1}{\Gamma(a)} \Gamma(a, x), \quad (1.2)$$

where we assume that a and x are positive.

The ratios $P(a, x), Q(a, x)$ are the standard chi-squared probability functions $P(\chi^2|\nu)$ and $Q(\chi^2|\nu)$ with parameters $a = \nu/2$ and $x = \chi^2/2$.

Not only the direct computation but also the inversion of cumulative distribution functions is an important topic in statistics, probability theory, communication theory and econometrics, in particular for computing percentage points of the gamma and beta distributions, which have the chi-square, F , and Student's t -distributions as special forms. In the tails of these distributions the numerical inversion is not very easy, and for these standard distributions asymptotic formulas are available.

As an example of application in Telecom Engineering, consider a communication system with one transmit and r receive antennas operating over a flat Rayleigh fading MIMO channel. For a given communication rate Φ , the outage probability can be expressed as [6]

* This work was supported by *Ministerio de Ciencia e Innovación*, project MTM2009-11686. NMT acknowledges financial support from the *Gobierno de Aragón*.

[†] Departamento de Matemática Aplicada y CC. de la Computación. ETSI Caminos. Universidad de Cantabria. 39005-Santander, Spain. (amparo.gil@unican.es).

[‡] Departamento de Matemáticas, Estadística y Computación. Facultad de Ciencias. Universidad de Cantabria. 39005-Santander, Spain. (javier.segura@unican.es).

[§] CWI, Science Park 123, 1098 XG Amsterdam, The Netherlands. (nico.temme@cwi.nl).

$$P_{out}^{(1 \times r)} = \frac{1}{\Gamma(r)} \gamma \left(r, \frac{2^\Phi - 1}{S/N} \right) \quad (1.3)$$

where S and N are the signal and noise power at the detection moment, respectively, and Γ and γ are the gamma and incomplete gamma functions.

On the other hand, for a communication system with t transmit and one receive antennas one has that the outage probability is given by [6]:

$$P_{out}^{(t \times 1)} = \frac{1}{\Gamma(t)} \gamma \left(t, \frac{2^\Phi - 1}{S/(tN)} \right). \quad (1.4)$$

Then, in order to express the communication rate Φ in terms of the desired outage probability P_{out} , one has to invert the incomplete gamma function.

In this paper, we present numerical algorithms for the computation and inversion of the incomplete gamma function ratios. A Fortran 90 version of the algorithms is made available at our website ¹.

In the numerical algorithms for the computation of the ratios, both $P(a, x)$ and $Q(a, x)$ are computed. First the primary function (the smaller of the two) is computed, and next the other one by using

$$P(a, x) + Q(a, x) = 1. \quad (1.5)$$

In particular for large values of a, x we have a transition at $a \sim x$, with

$$\begin{aligned} P(a, x) &\lesssim \frac{1}{2} \quad \text{when} \quad a \gtrsim x, \\ Q(a, x) &\lesssim \frac{1}{2} \quad \text{when} \quad a \lesssim x. \end{aligned} \quad (1.6)$$

In the next section we use more refined relations for small values of the parameters.

The algorithms are partly based on [3], where also negative values of a are considered for the pair $\{\Gamma(a, z), x^{-a}P(a, x)\}$, the second element being analytic at $x = 0$. For applications in probability theory and mathematical statistics we prefer working with the ratios. Also, when $a \sim x$ (both large) the second element becomes approximately $\frac{1}{2}x^{-a}$, and underflow may occur, in which case $P(a, x)$ cannot be computed. The element $\Gamma(a, x)$ may also become too large to handle.

For the large parameter case, in particular when $a \sim x$, we do not use Gautschi's approach (continued fractions), but use the method given in [8]; see also [4, §8.3] and [10, §5.2]. This method is based on uniform asymptotic expansions of the incomplete gamma functions, see [7]. In [2] these uniform expansions are also used. These authors use several expansions of coefficients for the critical region $a \sim x$, and we will explain that a more efficient expansion can be used.

For the numerical inversion we consider the equations

$$P(a, x) = p, \quad Q(a, x) = q, \quad 0 < p, q < 1, \quad (1.7)$$

for a given value of a and we compute x . For several cases we give new initial estimates for starting a numerical inversion process. The approach for large a is based on asymptotic methods developed in [9]; see also [4, §10.3.1] and [10, §6].

The resulting algorithms and software significantly improve both the accuracy and ranges of computation of the algorithm presented in [2].

¹<http://personales.unican.es/gila/incgam.zip>

2. Methods of computation. We describe the methods and the domains in the (x, a) quarter plane where they are used. First we define a function for separating the (a, x) quarter plane in two parts for assigning the primary function, that is, the function $P(a, x)$ or $Q(a, x)$ that has to be computed first.

As a minor modification of the function introduced in [3] we use

$$\alpha(x) = \begin{cases} x & \text{if } x \geq \frac{1}{2}, \\ \frac{\ln \frac{1}{2}}{\ln(\frac{1}{2}x)} & \text{if } 0 < x < \frac{1}{2}. \end{cases} \quad (2.1)$$

Then, the primary function is

$$\begin{aligned} P(a, x) & \quad \text{when } a \geq \alpha(x), \\ Q(a, x) & \quad \text{when } a < \alpha(x). \end{aligned} \quad (2.2)$$

The function $\alpha(x)$ has the same asymptotic behavior as the one given in [3] for small x . The changeover point $x = \frac{1}{2}$ is slightly better. At $x = a = \frac{1}{4}$ (changeover point in [3]) we have $P(a, x) \simeq 0.74$, while at the point $x = a = \frac{1}{2}$ we have $P(a, x) \simeq 0.68$.

In many representations we have to deal with the function

$$D(a, x) = \frac{x^a e^{-x}}{\Gamma(a+1)}, \quad (2.3)$$

see for example representation (2.11). For large values of a straightforward computation of $\Gamma(a+1)$ will cause overflow. When x is also large, say $x \sim a$, $D(a, x)$ may be computable, as can be seen when using Stirling's formula for the gamma function.

We write $D(a, x)$ in the form

$$D(a, x) = \frac{e^{-\frac{1}{2}a\eta^2}}{\sqrt{2\pi a} \Gamma^*(a)}, \quad (2.4)$$

where

$$\Gamma^*(a) = \frac{\Gamma(a)}{\sqrt{2\pi/a} a^a e^{-a}}, \quad a > 0, \quad (2.5)$$

and

$$\frac{1}{2}\eta^2 = \lambda - 1 - \ln \lambda, \quad \lambda = x/a. \quad (2.6)$$

The quantity η arises in other cases in this paper, and when we take the square root in (2.6) we assume that for $\lambda > 0$: $\text{sign}(\eta) = \text{sign}(\lambda - 1)$. Then, we have $\eta \sim \lambda - 1$ if $\lambda \sim 1$.

The function $\Gamma^*(a)$ has the asymptotic expansion (Stirling series)

$$\Gamma^*(a) \sim 1 + \frac{1}{12}a^{-1} + \frac{1}{288}a^{-2} + \dots, \quad a \rightarrow \infty. \quad (2.7)$$

This function is included in our software package.

For testing the algorithms it is useful to use the recurrence relations

$$P(a+1, x) = P(a, x) - D(a, x), \quad Q(a+1, x) = Q(a, x) + D(a, x). \quad (2.8)$$

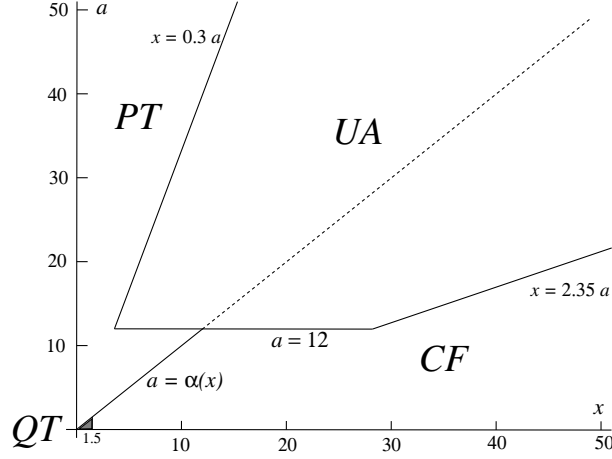


FIG. 2.1. The four domains for computing $P(a, x)$ and $Q(a, x)$.

For large values of a and x we can use a scaled version by writing

$$p(a, x) = P(a, x)/D(a, x), \quad q(a, x) = Q(a, x)/D(a, x), \quad x > 0, \quad (2.9)$$

and these functions satisfy the recursion

$$\frac{x}{a+1}p(a+1, x) = p(a, x) - 1, \quad \frac{x}{a+1}q(a+1, x) = q(a, x) + 1. \quad (2.10)$$

2.1. Domains of computation. In Figure 2.1 we indicate four domains of computation for the incomplete gamma function ratios corresponding to different methods:

PT: the Taylor expansion of $P(a, x)$, see §2.2.

QT: the Taylor expansion of $Q(a, x)$, see §2.3.

CF: the continued fraction for $Q(a, x)$, see §2.4.

UA: uniform asymptotic methods for $P(a, x)$ and $Q(a, x)$, see §2.5.

The domains of computation are established following a compromise between efficiency and accuracy: when two methods provide the same accuracy in a certain parameter region, the selection of one method or another will depend on the efficiency of each of these methods. The recurrence relations (2.8) and (2.10) will provide numerical checks for testing the accuracy of the resulting algorithm in all regions of the (x, a) -plane.

2.2. $P(a, x)$: Taylor expansion. The domain of computation of this section is indicated by **PT** in Figure 2.1.

The expansion is

$$P(a, x) = \frac{x^a e^{-x}}{\Gamma(a+1)} \sum_{n=0}^{\infty} \frac{x^n}{(a+1)_n}, \quad (2.11)$$

where we use the Pochhammer symbol $(a+1)_n = \Gamma(a+n+1)/\Gamma(a+1)$.

TABLE 2.1
The smallest number n satisfying (2.14), $\varepsilon = 10^{-15}$.

x a	$a/10$	$2a/10$	$3a/10$	$4a/10$	$5a/10$	$6a/10$	$7a/10$	$8a/10$
10	12	15	18	21	23	26	28	31
100	14	20	26	32	39	47	57	67
500	14	21	28	36	46	60	79	106
1000	14	21	28	36	48	63	86	122
5000	14	21	28	37	49	66	94	145
10000	14	21	28	37	49	67	95	149

The series converges for all a and x , and the rate of convergence improves as $a/x \rightarrow \infty$. The terms of the series are decreasing, because we apply this expansion when $a > x$.

To have an idea about the number of terms needed for obtaining a certain error ε after truncating the series we write

$$\sum_{n=0}^{\infty} \frac{x^n}{(a+1)_n} = S_{n_0}(a, x) + R_{n_0}(a, x), \quad (2.12)$$

where

$$S_{n_0}(a, x) = \sum_{n=0}^{n_0-1} \frac{x^n}{(a+1)_n}, \quad R_{n_0}(a, x) = \sum_{n=n_0}^{\infty} \frac{x^n}{(a+1)_n}, \quad (2.13)$$

and we compute the smallest $n = n_0$ that satisfies

$$\frac{x^n}{(a+1)_n} \leq \varepsilon. \quad (2.14)$$

In Table 2.1 we show the smallest number n satisfying (2.14) for $\varepsilon = 10^{-15}$ and several values of a and x . For large a we see that the number of terms needed (n) becomes constant. This can be understood by observing that the left-hand side of (2.14) becomes roughly λ^n , $\lambda = x/a$, if $a \gg n$. Because for large values of a we use the method of this section for $P(a, x)$ only if $\lambda \leq 3/10$, see Figure 2.1, it follows that not more than 30 terms are needed for $a \leq 10000$.

The remainder R_{n_0} in (2.13) can be written in the form

$$R_{n_0}(a, x) = \frac{x^{n_0}}{(a+1)_{n_0}} (a+n_0) e^x x^{-a-n_0} \int_0^x t^{a+n_0-1} e^{-t} dt, \quad (2.15)$$

which is again an incomplete gamma function. The function $u(t) = t - (a+n_0-1) \ln t$ is monotonic in $(0, x]$, and we can integrate with respect to u , giving

$$\int_0^x t^{a+n_0-1} e^{-t} dt = \int_{u(x)}^{\infty} e^{-u} f(u) du, \quad (2.16)$$

where

$$f(u) = -\frac{dt}{du} = \frac{t}{a+n_0-t-1} \leq \frac{x}{a+n_0-x-1}. \quad (2.17)$$

This gives

$$R_{n_0}(a, x) \leq \frac{x^{n_0}}{(a+1)_{n_0}} \frac{a+n_0}{a+n_0-x-1}. \quad (2.18)$$

2.3. $Q(a, x)$: Taylor expansion. The expansions of this section will be used for

$$0 \leq x \leq 1.5, \quad 0 \leq a \leq \alpha(x), \quad (2.19)$$

where $\alpha(x)$ is defined in (2.1). This domain is indicated by **QT** in Figure 2.1. For details and discussion we refer to [3, §4.1].

For this case we use the expansion

$$P(a, x) = \frac{x^a}{\Gamma(a)} \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{(a+n)n!}. \quad (2.20)$$

Straightforward use of the relation $Q(a, x) = 1 - P(a, x)$ should be avoided when a is small, and we write

$$Q(a, x) = u + v, \quad (2.21)$$

where

$$u = 1 - \frac{1}{\Gamma(1+a)} + \frac{1-x^a}{\Gamma(1+a)}. \quad (2.22)$$

For the first term we have available an algorithm to compute the function $g(a)$ in the representation

$$1 - \frac{1}{\Gamma(1+a)} = a(1-a)g(a). \quad (2.23)$$

The second term can be computed by using an expansion of $1 - x^a = 1 - \exp(a \ln(x))$.

The term v in (2.21) is

$$v = \frac{x^a}{\Gamma(1+a)} (1 - \Gamma(1+a)x^{-a}P(a, x)), \quad (2.24)$$

in which we can use the series in (2.20) by skipping the term with $n = 0$.

2.4. $Q(a, x)$: continued fraction. The domain of computation of this section is indicated by **CF** in Figure 2.1.

The continued fraction for $Q(a, x)$ is in the form [3]

$$Q(a, x) = \frac{x^a e^{-x}}{(x+1-a)\Gamma(a)} \left(\frac{1}{1+} \frac{a_1}{1+} \frac{a_2}{1+} \frac{a_3}{1+} \frac{a_4}{1+} \dots \right), \quad (2.25)$$

where

$$a_k = \frac{k(a-k)}{(x+2k-1-a)(x+2k+1-a)}, \quad k \geq 1. \quad (2.26)$$

The front factor term $(x+1-a)$ is not causing problems, because we use the continued fraction for $x \geq a$. When $a = 1, 2, 3, \dots$ the fraction is terminating.

TABLE 2.2

The smallest number n satisfying $t_n/S(a, x) < 10^{-15}$ for the series in (2.27), for some x and $a = x/2.35\rho$, with several values of ρ .

ρ	0.1	0.2	0.4	0.6	0.8	1.0	2.0	2.34
x								
1.5	58	58	57	56	55	54	51	52
2	45	45	44	43	42	40	40	32
10	13	12	12	11	11	11	14	14
100	5	6	7	8	9	11	27	39
500	4	5	6	6	7	8	25	68
1000	4	5	5	6	6	7	19	84
5000	4	4	4	5	5	5	10	133
10000	3	4	4	4	4	5	9	154

Several algorithms are available for the numerical evaluation of continued fractions. See, for example, [4, §6.6]. Gautschi used a conversion into an infinite series of the form

$$Q(a, x) = \frac{x^a e^{-x}}{(x+1-a)\Gamma(a)} S(a, x), \quad S(a, x) = 1 + \sum_{k=1}^{\infty} t_k, \quad (2.27)$$

where (with $\rho_0 = 0$)

$$t_k = \rho_1 \rho_2 \cdots \rho_k, \quad \rho_k = -\frac{a_k(1 + \rho_{k-1})}{1 + a_k(1 + \rho_{k-1})}, \quad k \geq 1. \quad (2.28)$$

In Table 2.2 we show the smallest number n satisfying $t_n/S(a, x) < 10^{-15}$ for several values of x and $a = x/2.35\rho$; for ρ see the table. The n -values in this table, as well as those of Table 2.1, are not a priori computed but obtained from numerical computations for the examples in the tables. When $\rho = 1$ and $a \geq 12$, the relation $a = x/2.35\rho$ corresponds with the border between the domains **UA** and **CF** in Figure 2.1. For other values of ρ and $x \geq 1.5$ the half-lines $a = x/2.35\rho$ are in the domain **CF**. Near the diagonal $x = a$ convergence becomes rather slow.

The continued fraction is an excellent alternative for the asymptotic expansion

$$Q(a, x) \sim \frac{x^{a-1} e^{-x}}{\Gamma(a)} \sum_{n=0}^{\infty} \frac{(-1)^n (1-a)_n}{x^n}, \quad x \rightarrow \infty. \quad (2.29)$$

2.5. $P(a, x), Q(a, x)$: uniform asymptotic expansion. The domain of computation of this section is indicated by **UA** in Figure 2.1. For more details on the used method we refer to [8]; see also [4, §8.3] and [10, §5.2]. We summarize the main steps for constructing an algorithm for this case.

We use the representations [7]

$$\begin{aligned} Q(a, x) &= \frac{1}{2} \operatorname{erfc}(\eta \sqrt{a/2}) + R_a(\eta), \\ P(a, x) &= \frac{1}{2} \operatorname{erfc}(-\eta \sqrt{a/2}) - R_a(\eta), \end{aligned} \quad (2.30)$$

where

$$\operatorname{erfc} x = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt, \quad (2.31)$$

the complementary error function. The quantity η is defined in (2.6), with again $\lambda = x/a$.

For $R_a(\eta)$ we have

$$R_a(\eta) = \frac{e^{-\frac{1}{2}a\eta^2}}{\sqrt{2\pi a}} S_a(\eta), \quad S_a(\eta) \sim \sum_{n=0}^{\infty} \frac{C_n(\eta)}{a^n}, \quad (2.32)$$

as $a \rightarrow \infty$. Note that the symmetry relation $P(a, x) + Q(a, x) = 1$ is preserved in the representations in (2.30) because $\operatorname{erfc} z + \operatorname{erfc}(-z) = 2$.

Although analytical expressions for the coefficients $C_n(\eta)$ are available, these representations are difficult to evaluate numerically for small values of η , that is near the transition $x \sim a$. In [2] power series expansions of the coefficients $C_n(\eta)$ for $n = 0, 1, 2, \dots, 9$ for small values of η are used.

For the present numerical algorithm we use a different approach. Instead of expanding each coefficient $C_n(\eta)$ we expand the function $S_a(\eta)$ of (2.32) in powers of η . The coefficients are functions of a , and we write

$$S_a(\eta) = \sum_{n=0}^{\infty} \alpha_n \eta^n. \quad (2.33)$$

To compute the coefficients α_n , we write

$$\alpha_n = \frac{\beta_n}{\Gamma^*(a)}, \quad n = 0, 1, 2, \dots, \quad (2.34)$$

where $\Gamma^*(a)$ is defined in (2.5). For β_n we have the recursion

$$\beta_n = \frac{1}{a}(n+2)\beta_{n+2} + d_{n+1}, \quad n = 0, 1, 2, \dots \quad (2.35)$$

where the coefficients d_n are defined as the coefficients in the power series

$$\frac{\eta}{\lambda - 1} = \sum_{n=0}^{\infty} d_n \eta^n. \quad (2.36)$$

This series and the one in (2.33) converge for $|\eta| < 2\sqrt{\pi}$.

The first values of d_n are

$$d_0 = 1, \quad d_1 = -\frac{1}{3}, \quad d_2 = -\frac{1}{12}, \quad d_3 = -\frac{2}{135}, \quad d_4 = \frac{1}{864}, \quad d_5 = \frac{1}{2835}. \quad (2.37)$$

To describe the algorithm, we choose a positive integer N , put $\beta_{N+2} = \beta_{N+1} = 0$, and compute the sequence

$$\beta_N, \beta_{N-1}, \dots, \beta_1, \beta_0 \quad (2.38)$$

from the recurrence relation (2.35). This recursion is stable in the backward direction.

Because

$$\alpha_1 = \frac{a}{\Gamma^*(a)} (\Gamma^*(a) - 1), \quad \Gamma^*(a) = 1 + \frac{1}{a} \beta_1, \quad (2.39)$$

we have

$$S_a(\eta) \approx \frac{a}{a + \beta_1} \sum_{n=0}^N \beta_n \eta^n \quad (2.40)$$

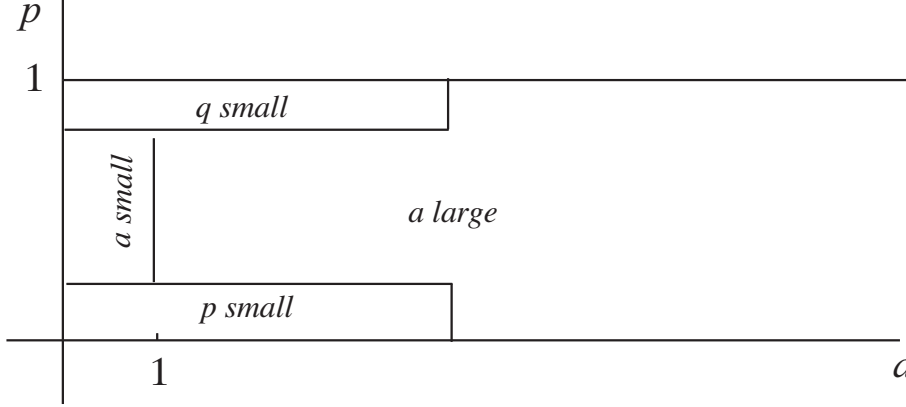


FIG. 3.1. The four domains for inverting $P(a, x) = p$ and $Q(a, x) = q$ with $p + q = 1$.

as an approximation for $S_a(\eta)$.

We use the approximation in (2.40) for computing the incomplete gamma functions in IEEE double precision for $a \geq 12$ and $|\eta| \leq 1$. We need the storage of 25 coefficients d_n , and in the series in (2.40). For $a = 12$ we need 25 terms; as a increases the convergence in the algorithm improves and we need a fewer number of terms.

The value $\eta = -1$ corresponds to $\lambda = 0.30\dots$, and the value $\eta = 1$ to $\lambda = 2.35\dots$. In Figure 2.1 we show the area indicated by **UA** in the (x, a) quarter-plane where we can apply the algorithm to obtain IEEE double precision.

3. Inversion methods. Several approaches are available in the (statistical) literature for computing the inverse of cumulative distribution functions, where often a first approximation of x is constructed, based on asymptotic estimates, but this first approximation may not be reliable. Higher approximations can be obtained by numerical inversion techniques, which require evaluation of the incomplete gamma functions. This may be rather time consuming, especially when a is large.

In [2] the inversion is considered also; we are using different methods based on analytic inversion of power series and asymptotic expansions. In this way it is clear how the first steps in the inversion method are taken. We use Newton methods when reliable starting values are available.

We solve the equations

$$P(a, x) = p, \quad Q(a, x) = q, \quad 0 \leq p \leq 1, \quad 0 \leq q \leq 1, \quad (3.1)$$

for x , with a as a given positive parameter. We consider several cases, which are schematically indicated in Figure 3.1. In most cases we invert the equation with $\min(p, q)$. Of course, if $x(p, a)$ denotes the solution of the first equation then the solution of the second equation satisfies $x(q, a) = x(1 - p, a)$. We assume that the user provides both p and q , which is important when $\min(p, q)$ is small.

3.1. Small values of p . When p is small we use the series in (2.20), and write the inversion problem as

$$x = r \left(1 + \sum_{n=1}^{\infty} \frac{a(-1)^n x^n}{(a+n)n!} \right)^{-1/a}, \quad r = (p\Gamma(1+a))^{1/a}, \quad (3.2)$$

where we assume that r is small. Inverting this relation, we obtain $x = r + \sum_{n=2}^{\infty} c_n r^n$, and the first few coefficients are

$$\begin{aligned} c_2 &= \frac{1}{a+1}, \\ c_3 &= \frac{3a+5}{2(a+1)^2(a+2)}, \\ c_4 &= \frac{8a^2+33a+31}{3(a+1)^3(a+2)(a+3)}, \\ c_5 &= \frac{125a^4+1179a^3+3971a^2+5661a+2888}{24(1+a)^4(a+2)^2(a+3)(a+4)}. \end{aligned} \quad (3.3)$$

It appears that $c_k = \mathcal{O}((a+1)^{-k+1})$ for large values of a , and from numerical experiments by checking several values of p, q, a which were relevant for these cases, we conclude that if $r < 0.2(1+a)$, that is, $p < (0.2(1+a))^a/\Gamma(1+a)$ we can obtain 4 digits accuracy in x with the coefficients shown in (3.3). This is enough for starting a Newton method for obtaining higher accuracy.

This method also works when a is small, because in that case $r \sim \exp(\frac{1}{a} \ln p)$ becomes small for all fixed $p \in (0, 1)$.

3.2. Small values of q . When q is small we use the asymptotic expansion in (2.29). A first approximation x_0 of x is obtained from the equation

$$e^{-x_0} x_0^a = q \Gamma(a) \quad (3.4)$$

Higher approximations of x are obtained in the form $x \sim x_0 - L + b \sum_{k=1}^{\infty} d_k / x_0^k$, where $b = 1 - a$, $L = \ln(x_0)$, with first coefficients

$$\begin{aligned} d_1 &= L - 1, \\ d_2 &= \frac{1}{2}(3b - 2bL + L^2 - 2L + 2), \\ d_3 &= \frac{1}{6}(24bL - 11b^2 - 24b - 6L^2 + \\ &\quad 12L - 12 - 9bL^2 + 6b^2L + 2L^3), \\ d_4 &= \frac{1}{12}(72 + 36L^2 + 3L^4 - 72L + 162b - 168bL - 12L^3 + \\ &\quad 25b^3 - 22bL^3 + 36b^2L^2 - 12b^3L + 84bL^2 + 120b^2 - 114b^2L). \end{aligned} \quad (3.5)$$

These coefficients, as well as those given in (3.3), are obtained by symbolic computation. Maple codes for computing these or more coefficients can be obtained at our website ².

This method works with rather small values of q (large values of x_0). When we assume that $0 < a < 10$ and $q < e^{-\frac{1}{2}a}/\Gamma(a+1)$ we obtain about 4 digits accuracy in x , which is enough for starting a Newton method for obtaining higher accuracy. For larger values of a and q , the method of §3.4 can be used.

²<http://personales.unican.es/gila/coefMaple.txt>

3.3. Small values of a . We consider the inversion of $P(a, x) = p$ for $a \in (0, 1)$. For these values of a we cannot derive expansions for obtaining a reliable starting value, unless p is small in which case we can use the result of §3.1. Also, the method of the next section cannot be used in this case.

We observe that

$$P(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt < \frac{1}{\Gamma(a)} \int_0^x t^{a-1} dt = \frac{x^a}{\Gamma(a+1)} \quad (3.6)$$

and

$$P(a, x) > \frac{1}{\Gamma(a)} \int_0^x e^{-t} dt = \frac{1}{\Gamma(a)} (1 - e^{-x}). \quad (3.7)$$

Let x_l, x_u be defined by

$$x_l = (p\Gamma(a+1))^{1/a}, \quad x_u = -\ln(1 - p\Gamma(a+1)). \quad (3.8)$$

Then the solution x of $P(a, x) = p$ with $a \in (0, 1)$ satisfies $x_l < x < x_u$. The same results hold for the inversion of $Q(a, x) = q$ for $a \in (0, 1)$ when p is replaced with $1 - q$.

These bounds of x can be used for starting values for the Newton method. Of the two possibilities, the best option is x_l because the Newton method necessarily converges from this starting value. The reason is that $P(a, x)$ is an increasing function with negative second derivative; elementary graphical arguments show that if an starting value $x_0 < x$ is chosen, then the Newton iteration $x_{n+1} = x_n - P(a, x_n)/P'(a, x_n)$ produces a monotonically increasing sequence which is bounded by x , and therefore a converging sequence. The same is true for $Q(a, x)$ because it is decreasing and with positive second derivative.

For this case, the initial approximation may be inaccurate, however convergence is certain and not very expensive, as numerical experiments show.

3.4. Large values of a . We perform the inversion of (3.1) with respect to the parameter η by using the representations (2.30). Afterwards we have to compute λ and x from the relation for η in (2.6) and $\lambda = x/a$. We concentrate on the second equation in (3.1). For details we refer to [9]; see also [4, §10.3.1] and [10, §6].

We rewrite the inversion problem in the form

$$\frac{1}{2}\operatorname{erfc}\left(\eta\sqrt{a/2}\right) + R_a(\eta) = q, \quad q \in [0, 1], \quad (3.9)$$

which is equivalent to the second equation in (3.1), and we denote the solution of the above equation by $\eta(q, a)$.

To start the procedure we consider $R_a(\eta)$ in (3.9) as a perturbation, and we define the number $\eta_0 = \eta_0(q, a)$ as the real number that satisfies the equation

$$\frac{1}{2}\operatorname{erfc}\left(\eta_0\sqrt{a/2}\right) = q. \quad (3.10)$$

Computation of η_0 requires an inversion of the complementary error function, which is discussed in §3.4.1.

For large values of a the value η defined by (3.9) can be approximated by η_0 :

$$\eta(q, a) = \eta_0(q, a) + \varepsilon(\eta_0, a), \quad (3.11)$$

and it is possible to expand

$$\varepsilon(\eta_0, a) \sim \frac{\varepsilon_1(\eta_0, a)}{a} + \frac{\varepsilon_2(\eta_0, a)}{a^2} + \frac{\varepsilon_3(\eta_0, a)}{a^3} + \dots, \quad (3.12)$$

as $a \rightarrow \infty$. The coefficients $\varepsilon_j(\eta_0, a)$ can be written explicitly as functions of η_0 , the first coefficient being

$$\varepsilon_1(\eta_0, a) = \frac{1}{\eta_0} \ln \frac{\eta_0}{\lambda_0 - 1}, \quad (3.13)$$

where λ_0 follows from inverting (2.6) with η replaced by η_0 .

From numerical tests it follows that we can obtain 3 or 4 significant digits when using (3.12) with 4 terms for $a \geq 1$ and $q \in (0, 1)$. This is enough to start a stable Newton method.

REMARK 1. We start the inversion of $P(a, x) = p$ with the equation

$$\frac{1}{2} \operatorname{erfc} \left(-\eta \sqrt{a/2} \right) - R_a(\eta) = p, \quad p \in [0, 1], \quad (3.14)$$

and with the first approximation $\eta_0(p, a) = -\eta_0(q, a)$ (cf. (3.10)), with $p + q = 1$. The results for $P(a, x) = p$ then follow from the results of this section with $\eta_0(q, a)$ replaced by $-\eta_0(p, a)$, throughout.

3.4.1. Complementary error function. The inversion of this function is the first step in the large a asymptotic inversion method for the incomplete gamma function ratios. We summarize results from [11, §7.17]; for more details, see [4, §10.2].

We denote the inverse of the function $x = \operatorname{erfc} y$ by $y = \operatorname{inverfc} x$. Then, with $t = \frac{1}{2} \sqrt{\pi} (1 - x)$, we have

$$\operatorname{inverfc} x = t + \frac{1}{3} t^3 + \frac{7}{30} t^5 + \frac{127}{630} t^7 + \dots, \quad 0 < x < 2. \quad (3.15)$$

This expansion is actually for the inversion of $\operatorname{erf} y = x$, useful for small values of t . More coefficients in the expansion can be found in [5], but they can easily be obtained by using computer algebra and formal manipulation of power series.

For small values of x we have an asymptotic expansion. Let t , α , and β be defined by

$$t = \frac{2}{\pi x^2}, \quad \alpha = \frac{1}{\ln t}, \quad \beta = \ln(\ln t). \quad (3.16)$$

Then we have the expansion

$$\operatorname{inverfc} x \sim \frac{1}{\sqrt{2\alpha}} (1 + x_1 \alpha + x_2 \alpha^2 + x_3 \alpha^3 + x_4 \alpha^4 + \dots). \quad (3.17)$$

The first coefficients x_k are given by

$$\begin{aligned} x_1 &= -\frac{1}{2} \beta, \\ x_2 &= -\frac{1}{8} (\beta^2 - 4\beta + 8), \\ x_3 &= -\frac{1}{16} (\beta^3 - 8\beta^2 + 32\beta - 56), \\ x_4 &= -\frac{1}{384} (15\beta^4 - 184\beta^3 + 1152\beta^2 - 4128\beta + 7040). \end{aligned} \quad (3.18)$$

In §4.1 we use a higher-order Newton process to obtain better approximations.

For other methods of the inversion of the error functions, we refer to [5], where coefficients of the Maclaurin expansion of $y = \operatorname{inverf} x$, the inverse of $x = \operatorname{erf} y$, are given, with Chebyshev coefficients for an expansion on the y -interval $[-0.8, 0.8]$. For small values of x (not smaller than 10^{-300}) high-precision coefficients of Chebyshev expansions are given for the numerical evaluation of $y = \operatorname{inverfc} x$. For rational Chebyshev (near-minimax) approximations for $y = \operatorname{inverfc} x$, we refer to [1], where x -values are considered in the x -interval $[10^{-10000}, 1]$, with relative errors ranging down to 10^{-23} . An asymptotic formula for the region $x \rightarrow 0$ is also given.

4. High order Newton-like methods. We describe the known method (see [4, §10.6] and [10, §6.8]) for constructing Newton-like methods of high order, with details for the inversion of the complementary error function and the incomplete gamma functions. For certain values of the parameters we will use these methods in the inversion algorithms.

The method is based on inverting the power series in h :

$$f(\zeta) = f(\zeta_0 + h) = f(\zeta_0) + hf_1 + \frac{1}{2!}h^2f_2 + \frac{1}{3!}h^3f_3 + \cdots, \quad (4.1)$$

where ζ is a zero of f and ζ_0 is an approximation. Also, $f_k = f^{(k)}(\zeta_0)$. We expand, assuming $f(\zeta_0)$ is small,

$$h = c_1f(\zeta_0) + c_2f^2(\zeta_0) + c_3f^3(\zeta_0) + \cdots. \quad (4.2)$$

and find, when $f_1 \neq 0$,

$$\begin{aligned} c_1 &= -\frac{1}{f_1}, & c_2 &= -\frac{f_2}{2f_1^3}, \\ c_3 &= \frac{-3f_2^2 + f_3f_1}{6f_1^5}, & c_4 &= -\frac{f_4f_1^2 + 15f_2^3 - 10f_2f_3f_1}{24f_1^7}. \end{aligned} \quad (4.3)$$

When we neglect in (4.2) the coefficients c_k with $k \geq 2$, we obtain Newton's Rule, with $\zeta \doteq \zeta_0 - f(\zeta_0)/f'(\zeta_0)$.

When $f(z)$ satisfies a simple ordinary differential equation, the higher derivatives can be replaced by combinations of lower derivatives.

4.1. Complementary error function. We compute a zero of the function $f(y) = \operatorname{erfc} y - x$ with $0 < x < 2$. Assume we have a first approximation y_0 then the zero y is written as $y = y_0 + h$, where h is as in (4.2) with $f(\zeta_0) = f(y_0)$. The derivatives of $f(y)$ are in terms of Hermite polynomials³:

$$f_k = ce^{-y^2}(-1)^{k-1}H_{k-1}(y), \quad k = 1, 2, 3, \dots, \quad c = -\frac{2}{\sqrt{\pi}}. \quad (4.4)$$

This gives

$$\begin{aligned} c_1 &= -1/f_1 = -e^{y_0^2}/c, & c_2 &= y_0c_1^2, & c_3 &= \frac{1}{3}(4y_0^2 + 1)c_1^3, \\ c_4 &= \frac{1}{6}y_0(7 + 12y_0^2)c_1^4, & c_5 &= \frac{1}{30}(8y_0^2 + 7)(12y_0^2 + 1)c_1^5. \end{aligned} \quad (4.5)$$

³<http://dlmf.nist.gov/18.5.E5>

4.2. Incomplete gamma function. When starting with an initial value $x_0 > 0$ for the inversion of $f(x) = P(a, x) - p$, the coefficients c_k can be derived from the integral representation of $P(a, x)$. This gives

$$\begin{aligned} c_1 &= -x_0^{1-a} e^{x_0} \Gamma(a), \quad c_2 = \frac{x_0 + 1 - a}{2x_0} c_1^2, \\ c_3 &= \frac{2x_0^2 + (a-1)(2a-1-4x_0)}{6x_0^2} c_1^3, \\ c_4 &= \frac{6x_0^3 - (a-1)(18x_0^2 - x_0(18a-11) + (2a-1)(3a-1))}{24x_0^3} c_1^4. \end{aligned} \quad (4.6)$$

For $a = \frac{1}{2}$, in which case $P(\frac{1}{2}, x) = \operatorname{erf} \sqrt{x}$, and $p = \frac{1}{2}$, we take x_0 of (3.8), giving $x_0 = \pi/16 = 0.19634954$ and $f(x_0) = \operatorname{erf} \sqrt{x_0} - \frac{1}{2} = -0.03088405$. Using 4 terms c_k from (4.6), it follows that $h = 0.03111855$, giving $x \doteq x_0 + h = 0.227468092$, and with this value we have $f(x) \doteq -1.12 \cdot 10^{-7}$.

For much smaller values of a , this fourth order method has some convergence problems. However, as commented before, the plain Newton method converges with certainty. For values of a smaller than 0.05 the plain Newton is used in the algorithm.

5. Associated Software and Testing. A Fortran 90 module (**IncgamFI**) implementing the algorithms can be obtained at our website⁴. The module includes the public routines **incgam**, for the computation of $P(a, x)$ and $Q(a, x)$, and **invincgam**, for the computation of x in the equations $P(a, x) = p$ and $Q(a, x) = q$ with a as a given positive parameter; p and q are also inputs of this routine. As a test of the routine **incgam**, we compute the maximum relative errors for the relations given in equation (2.8) using 10^6 and 10^7 random points in the following two regions of the (x, a) -plane, respectively. We obtain:

1. $(0, 1] \times (0, 1]$: $1.7 \cdot 10^{-15}$,
2. $(0, 500] \times (0, 500]$: $7.9 \cdot 10^{-13}$.

These errors constitute the accuracy claim of our algorithm **incgam**.

Complementarily, we have performed tests in larger ranges by using the scaled expressions given in (2.9) and considering the relation (2.10) in the regions where the function $D(a, x)$ (2.3) can be explicitly factored out. The use of scaled functions allows to test these methods for large values of x and a . We consider first the cases of the Taylor expansion and the continued fraction method. Using 10^7 and 10^8 random points in the following two regions of the (x, a) -plane (excluding the points where asymptotic expansions are used, see Figure 2.1), we obtain as maximum relative errors:

1. $(0, 10^4] \times (0, 10^4]$: $8.3 \cdot 10^{-15}$,
2. $(0, 10^5] \times (0, 10^5]$: $9.1 \cdot 10^{-15}$.

For testing the uniform asymptotic expansions for large values of the parameters, in the numerical algorithm we use the representations

$$\begin{aligned} Q(a, x) &= \Gamma^*(a) D(a, x) \left(\frac{1}{2} \sqrt{2\pi a} \operatorname{Erfc}(\eta \sqrt{a/2}) + S_a(\eta) \right), \\ P(a, x) &= \Gamma^*(a) D(a, x) \left(\frac{1}{2} \sqrt{2\pi a} \operatorname{Erfc}(-\eta \sqrt{a/2}) - S_a(\eta) \right), \end{aligned} \quad (5.1)$$

where $\operatorname{Erfc} x = e^{x^2} \operatorname{erfc} x$, $D(a, x)$ is defined in (2.3), and $S_a(\eta)$ denotes the series as in (2.32). In this way we can test the algorithm for the scaled function

⁴<http://personales.unican.es/gila/incgam.zip>

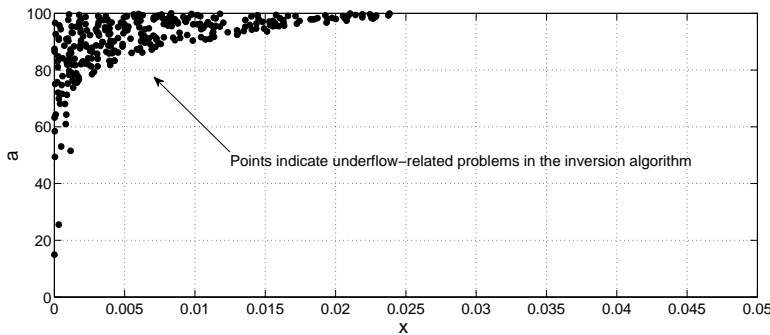


FIG. 5.1. Test of the inversion algorithm: underflow-related problems in double precision arithmetic in the region $(x, a) \in (0, 100) \times (0, 100)$. Notice that the problems are concentrated in the interval $0 < x < 0.025$.

$q(a, x) = Q(a, x)/D(a, x)$ by using the relation in (2.10). Similarly for $p(a, x) = Q(a, x)/D(a, x)$. Using 10^7 random points in the region $(0, 10^4] \times (0, 10^4]$ of the (x, a) -plane, the maximum relative error obtained when computing the scaled functions using uniform asymptotic expansions is $4 \cdot 10^{-14}$.

The results obtained with the scaled expressions confirm the stability and accuracy of the methods used for computing $P(a, x)$ and $Q(a, x)$.

For the inversion algorithm, testing is made by checking that the composition of the functions with their inverse is the identity: from the values of a and $x = x_{in}$ we compute $p = P(a, x_{in})$ and $q = Q(a, x_{in})$ and from the values of p, q and a we obtain x_{out} from the inversion algorithm; we compare the values x_{in} and x_{out} and compute the relative errors. As a first test, we generate 10^7 random points in the region $(x, a) \in (0, 100] \times (0, 100]$. In the test we exclude the points where problems related to the underflow limit in double precision arithmetic when computing the incomplete gamma function ratios appear (see Figure 5.1). The maximum relative error is $1.42 \cdot 10^{-11}$. It is important to note that this value is obtained for a point (x, a) near the region of underflow problems: the value of the function $P(a, x)$ at that point was $\sim 10^{-297}$. Some loss of accuracy is expected in these cases.

The accuracy of the initial estimates discussed in sections (3.1), (3.2) and (3.4) is illustrated in Figure 5.2, where relative distances between the initial estimate (x_{ini}) and the true value (x) are plotted. 10^4 random points have been considered in the plane $(x, a) \in (0, 100) \times (1, 100)$. As can be seen, the poorest estimate in the test is located at a relative distance less than $5 \cdot 10^{-3}$ to the real value.

The number of iterations used in the inversion algorithm is also tested. Figure 5.3 shows the number of iterations used in the region $(x, a) \in (0, 100] \times (0, 100]$ for computing the x values in the equations $p = P(a, x)$ and $q = Q(a, x)$ within an accuracy of $10^{-12} - 10^{-14}$. As can be seen, 2 or 3 iterations are enough in most of the points of (x, a) -plane.

As a final comment, our module **IncgamFI** clearly improves the algorithm provided in [2], having its associated Fortran 77 routines **GRATIO** and **GAMINV** (single precision routines) both a more limited range of validity and accuracy than our algorithms. As an illustration, Figure 5.4 shows the points where the inversion routine **GAMINV** fails when performing the same test as used for our inversion algorithm. A point is plotted when the relative accuracy was greater than 0.1. The

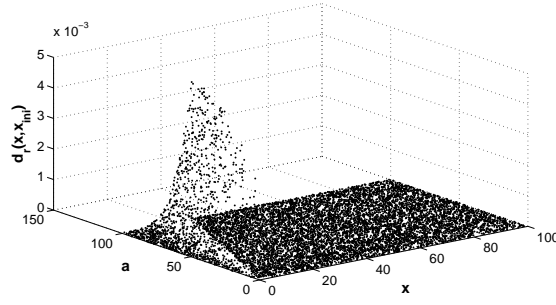


FIG. 5.2. Accuracy of the initial estimates for the inversion problem: points correspond to relative distances $d_r(x, x_{ini}) = \|1 - x_{ini}/x\|$ between the initial estimates (x_{ini}) and the true values x .

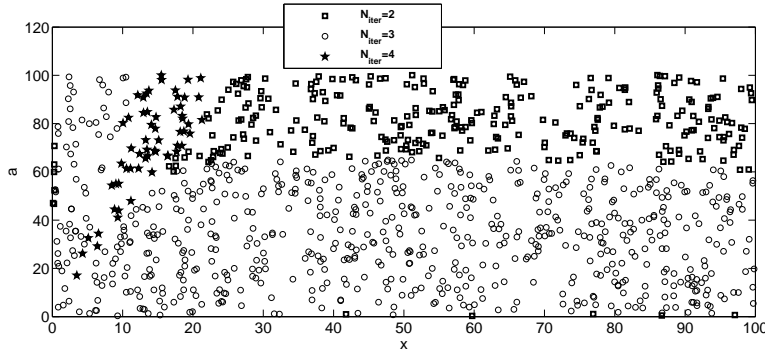


FIG. 5.3. Number of iterations used in the inversion algorithm in the region $(x, a) \in (0, 100] \times (0, 100]$.

routine **GRATIO** was used for the direct computation of the functions.

Acknowledgments. This work was supported by *Ministerio de Ciencia e Innovación*, project MTM2009-11686.

REFERENCES

- [1] J. M. Blair, C. A. Edwards, and J. H. Johnson. Rational Chebyshev approximations for the inverse of the error function. *Math. Comp.*, 30(136, loose microfiche suppl.):7–68, 1976.
- [2] A.R. DiDonato and A. H. Morris, Jr. Computation of the incomplete gamma function ratios and their inverse. *ACM Trans. Math. Software*, 12(4):377–393, December 1986.
- [3] W. Gautschi. A computational procedure for incomplete gamma functions. *ACM Trans. Math. Software*, 5(4):466–481, 1979.
- [4] A. Gil, J. Segura, and N. M. Temme. *Numerical methods for special functions*. SIAM, Philadelphia, PA, 2007.
- [5] A. Strecok. On the calculation of the inverse of the error function. *Math. Comp.*, 22:144–158, 1968.
- [6] E Telatar. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun.*, 10(6):585–595, 1999.
- [7] N. M. Temme. The asymptotic expansion of the incomplete gamma functions. *SIAM J. Math. Anal.*, 10(4):757–766, 1979.
- [8] N. M. Temme. On the computation of the incomplete gamma functions for large values of

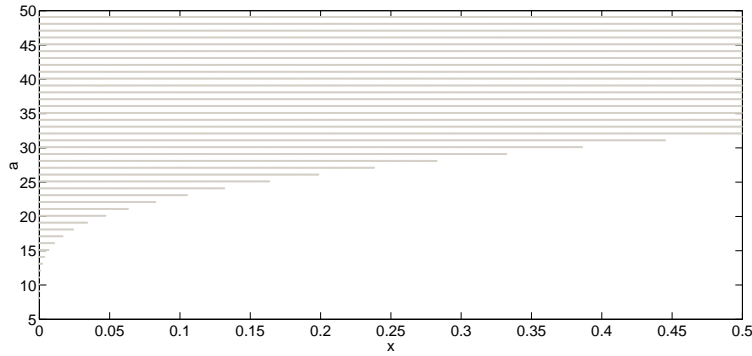


FIG. 5.4. Inversion routine **GAMINV** of reference [2]: points are plotted when the relative accuracy in the inversion test is greater than 0.1. The routine **GRATIO** was used for the direct computation of the functions.

the parameters. In *Algorithms for approximation* (Shrivenham, 1985), volume 10 of *Inst. Math. Appl. Conf. Ser. New Ser.*, pages 479–489. Oxford Univ. Press, New York, 1987.

- [9] N. M. Temme. Asymptotic inversion of incomplete gamma functions. *Math. Comp.*, 58(198):755–764, 1992.
- [10] N. M. Temme. Numerical aspects of special functions. *Acta Numer.*, 16:379–478, 2007.
- [11] N. M. Temme. Error functions, Dawson’s and Fresnel integrals. In *NIST handbook of mathematical functions*, pages 159–171. U.S. Dept. Commerce, Washington, DC, 2010. <http://dlmf.nist.gov/7>.